

INEQUALITIES FOR NONCENTRAL CHI-SQUARE DISTRIBUTIONS

Wilbert C.M. KALLENBERG

Faculty of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Received November 1988

Revised March 1989

Abstract: An upper and lower bound are presented for the difference between the distribution functions of noncentral chi-square variables with the same degrees of freedom and different noncentralities. The inequalities are applied in a comparison of two approximations to the power of Pearson's chi-square test.

Keywords: Noncentral χ^2 distribution, Pearson's chi-square test, contamination family, exponential family.

1. Main results

Noncentral chi-square distributions play an important role in statistics. They occur for instance as exact or limiting distributions of test statistics under alternatives. To examine the power functions of such tests we come across the problem of comparing distribution functions of noncentral chi-square variables. In this section inequalities are presented, which make such a comparison possible. They show how differences between noncentralities are carried over to differences between their distribution functions.

A nontrivial application is given in Section 2, where two approximations to the power of Pearson's chi-square test are investigated.

In the following theorem an *upper bound* is given for the difference between the distribution functions of noncentral chi-square variables with the same degrees of freedom and different noncentralities. For $k = 1, 2, \dots$, and $\delta \geq 0$ denote by $\chi_k^2(\delta)$ an r.v. with a noncentral chi-square distribution with k degrees of freedom and noncentrality parameter δ .

Theorem 1.1. *Let δ and δ^* be nonnegative real numbers. Then for all $c \in \mathbb{R}$ and $k \geq 1$,*

$$\left| \Pr(\chi_k^2(\delta) \leq c) - \Pr(\chi_k^2(\delta^*) \leq c) \right| \leq (2\pi)^{-1/2} |\delta^{1/2} - \delta^{*1/2}| \Pr(\chi_{k-1}^2(0) \leq c). \quad (1.1)$$

Moreover, if $\delta \leq \delta^*$, then for all $c \in \mathbb{R}$ and $k \geq 1$,

$$0 \leq \Pr(\chi_k^2(\delta) \leq c) - \Pr(\chi_k^2(\delta^*) \leq c) \leq (2\pi)^{-1/2} (\delta^{1/2} - \delta^{*1/2}) \Pr(\chi_{k-1}^2(0) \leq c). \quad (1.2)$$

Proof. Since (1.2) implies (1.1) it suffices to prove (1.2). So let $\delta \leq \delta^*$. Writing g for the density of $\chi_{k-1}^2(0)$, Φ for the distribution function of the $N(0, 1)$ -distribution and ϕ for its density we have

$$\begin{aligned} & \Pr(\chi_k^2(\delta) \leq c) - \Pr(\chi_k^2(\delta^*) \leq c) \\ &= \int_0^c \left\{ \left[\Phi(\delta^{1/2} + (c-v)^{1/2}) - \Phi(\delta^{1/2} - (c-v)^{1/2}) \right] \right. \\ & \quad \left. - \left[\Phi(\delta^{*1/2} + (c-v)^{1/2}) - \Phi(\delta^{*1/2} - (c-v)^{1/2}) \right] \right\} g(v) dv \\ &= \int_0^c \int_{\delta^{1/2} - (c-v)^{1/2}}^{\delta^{*1/2} - (c-v)^{1/2}} \phi(u) \left[1 - \exp \left\{ -2(c-v)^{1/2} (u + (c-v)^{1/2}) \right\} \right] du g(v) dv. \end{aligned} \quad (1.3)$$

Since

$$0 \leq \phi(u) [1 - \exp\{-2h(u+h)\}] \leq (2\pi)^{-1/2} \quad \text{for all } u, h \in \mathbb{R},$$

the proof of (1.2) is complete. \square

Remark 1.1. The constant $(2\pi)^{-1/2}$ in (1.1) and (1.2) is sharp. This is seen by taking $c = \delta$ and $\delta^{*1/2} = \delta^{1/2} + (\log \delta)^{-1}$ and noting that in that case

$$\lim_{\delta \rightarrow \infty} \{ \Pr(\chi_k^2(\delta) \leq c) - \Pr(\chi_k^2(\delta^*) \leq c) \} \{ \delta^{1/2} - \delta^{*1/2} \}^{-1} = (2\pi)^{-1/2}.$$

The lower bound can not be of the form $c|\delta^{1/2} - \delta^{*1/2}|$ for all δ and δ^* , because $c|\delta^{1/2} - \delta^{*1/2}| > 1$ if $|\delta^{1/2} - \delta^{*1/2}| > c^{-1}$. Therefore we assume that $\delta^{1/2} - \delta^{*1/2}$ is bounded. Moreover, if both δ and δ^* tend to 0, the lower bound is $o(|\delta^{1/2} - \delta^{*1/2}|)$. So to obtain the reverse of Theorem 1.1 we assume that $\max(\delta, \delta^*)$ is bounded away from 0.

Theorem 1.2. For each $n = 1, 2, \dots$, let δ_n and δ_n^* be nonnegative real numbers. If $\liminf_{n \rightarrow \infty} \max(\delta_n, \delta_n^*) > 0$ and $\delta_n^{1/2} - \delta_n^{*1/2} = O(1)$ as $n \rightarrow \infty$, then for each $k \geq 1$ there exists $c_1 > 0$ such that

$$\sup_{c > 0} |\Pr(\chi_k^2(\delta_n) \leq c) - \Pr(\chi_k^2(\delta_n^*) \leq c)| \geq c_1 |\delta_n^{1/2} - \delta_n^{*1/2}|. \quad (1.4)$$

Proof. Without loss of generality assume $\delta_n \leq \delta_n^*$. If $\liminf_{n \rightarrow \infty} \delta_n^* > 0$ and $\delta_n^{1/2} - \delta_n^{*1/2} = O(1)$, it is easily seen that for $c = \delta_n^* + 1$ and all v with $0 < v \leq \frac{1}{2}$,

$$\int_{\delta_n^{1/2} - (c-v)^{1/2}}^{\delta_n^{*1/2} - (c-v)^{1/2}} \phi(u) [1 - \exp\{-2(c-v)^{1/2}(u + (c-v)^{1/2})\}] du \geq c_2 |\delta_n^{1/2} - \delta_n^{*1/2}|$$

for some positive constant c_2 . In view of (1.3) the result now easily follows. \square

Remark 1.2. Theorems 1.1 and 1.2 state that the difference between the distribution functions of $\chi_k^2(\delta)$ and $\chi_k^2(\delta^*)$ is of order $\delta^{1/2} - \delta^{*1/2}$. On the other hand $E\chi_k^2(\delta) - E\chi_k^2(\delta^*) = \delta - \delta^*$, which is of larger order than $\delta^{1/2} - \delta^{*1/2}$ if $\delta, \delta^* \rightarrow \infty$. This seems to be a contradiction. However, for large δ the density at x of $\chi_k^2(\delta)$ is of order $\delta^{-1/2}$ for x in a neighbourhood of its expectation δ (cf. Drost et al., 1989a, Lemma 3.4). Therefore a 'shift' of order $\delta - \delta^*$ indeed corresponds to a difference of order $\delta^{1/2} - \delta^{*1/2}$ between the distribution functions, in the sense that in general $F(x + \delta - \delta^*) - F(x)$ is of the same order as $f(x)(\delta - \delta^*)$ if $f = F'$, which is of order $\delta^{1/2} - \delta^{*1/2}$ if f is of order $\delta^{-1/2}$.

2. Pearson's chi-square test

Let Y_1, \dots, Y_n be i.i.d. real valued random variables with an absolutely continuous distribution. Consider the problem of testing the simple hypothesis H_0 that the Y_j 's have given density h . The classical Pearson chi-square test may be employed if one aims at a broad range of alternatives. To study the power behaviour at any given alternative density g , a family $\{g_\theta: 0 \leq \theta \leq 1\}$ is introduced with $g_0 = h$ and $g_1 = g$. The general idea is that the power of a test at g , for moderate sample sizes will be reflected in the asymptotic power of the test for $n \rightarrow \infty$ and $\theta = \theta_n \rightarrow 0$. This approach has proved to be very effective in parametric models, here the main difficulty is that no single natural parametric family of densities 'between h and g ' exists. A favourite family is the so called *contamination* family, where

$$g_\theta = (1 - \theta)h + \theta g, \quad 0 \leq \theta \leq 1 \quad (2.1)$$

(cf. e.g. Gvanceladze and Chibisov, 1979; Quine and Robinson, 1985). In Kallenberg et al. (1985) also a so called *exponential* family is considered. In this case we write \tilde{g}_θ , given by

$$\tilde{g}_\theta = c(\theta)g^\theta h^{1-\theta}, \quad 0 \leq \theta \leq 1, \quad (2.2)$$

where $c(\theta)$ is a normalizing constant. Such a family occurs also in measuring distance between h and g in terms of Chernoff's index (cf. Brown, 1971; Kester and Kallenberg, 1986). The model (2.1) may be called arithmetic, while (2.2) is a geometric model.

Example 2.1. Let h be the density of a standard normal distribution and let g be the density of a normal $N(1, 1)$ -distribution. The exponential family between h and g is the family of normal $N(\theta, 1)$ -distributions with $0 \leq \theta \leq 1$. At first sight this seems to be a more natural family between h and g than the contamination family, where g_θ is *not normal* for $0 < \theta < 1$. In general, when g and h are members of some exponential family, \tilde{g}_θ belongs to the same exponential family for $0 \leq \theta \leq 1$.

In situations as in the preceding example one is inclined to predict the power of Pearson's chi-square test by using local theory in exponential families, since the family of normal $N(\theta, 1)$ -distributions with $0 \leq \theta \leq 1$ seems to be the natural parametric family between a standard normal distribution and a normal $N(1, 1)$ -distribution. On the other hand, the classical noncentral chi-square approximation to the power of Pearson's chi-square test corresponds to local theory based on contamination families as will be shown. In that vein the contamination family is the natural choice even in the case of Example 2.1.

Numerical results in Kallenberg et al. (1985) indicate that in general local theory in contamination families is more useful in predicting power behaviour of Pearson's chi-square test than local theory in exponential families. In this section a theoretical explanation is given for the numerical evidence by comparing the error terms of both approximations. The inequalities of Section 1 are the main tools in such a comparison.

Before stating the results we introduce some notation. In line with applications, we assume that $h > 0$ a.e. (λ) on a possibly infinite interval S and $h = 0$ outside S , where λ is the Lebesgue measure. It is also assumed that $g = 0$ outside S ; this involves no loss of generality, since goodness-of-fit tests are always performed conditional on the event that no observation lies outside S . Let A_1, \dots, A_k be a partition of S into nondegenerate disjoint intervals. We restrict ourselves to a *fixed* number of classes k . Define

$$Z_i = \# \{ Y_j \in A_i \} \quad \text{and} \quad \pi_i = \int_{A_i} h \, d\lambda.$$

Pearson's chi-square statistic is defined as

$$X^2 = \sum_{i=1}^k (Z_i - n\pi_i)^2 (n\pi_i)^{-1}.$$

For local alternatives $\{g_{\theta_n}\}$ or $\{\tilde{g}_{\theta_n}\}$ such that $\lim_{n \rightarrow \infty} n\theta_n^2$ exists finitely, the asymptotic distribution of X^2 as $n \rightarrow \infty$ is $\chi_{k-1}^2(\lim_{n \rightarrow \infty} n\theta_n^2 \Delta)$ in the contamination family and $\chi_{k-1}^2(\lim_{n \rightarrow \infty} n\theta_n^2 \tilde{\Delta})$ in the exponential family (cf. Lemma A.1 in the Appendix), where

$$\Delta = \Delta(g) = \sum_{i=1}^k \pi_i^{-1} \left(\int_{A_i} g \, d\lambda - \int_{A_i} h \, d\lambda \right)^2 \quad \text{and} \quad \tilde{\Delta} = \tilde{\Delta}(g) = \sum_{i=1}^k \pi_i^{-1} \left(\int_{A_i} fh \, d\lambda \right)^2 \quad (2.3)$$

with

$$f = \log(g/h) - \int h \log(g/h) \, d\lambda.$$

(It is assumed that $\int h \log(g/h) \, d\lambda$ is finite.)

In view of the general idea underlying families between h and g , the preceding limiting result is applied (e.g. in making inference on the power behaviour of tests for different values of k) as an approximation theorem in the following way. Substituting $\theta_n = 1$, the power of the X^2 -test for moderate sample sizes is calculated using $\chi_{k-1}^2(n \Delta(g))$ in the contamination family approach, or $\chi_{k-1}^2(n \tilde{\Delta}(g))$ in the exponential family case. Note that the contamination family approach leads to the traditional approximation to the power of the X^2 -test (cf. Serfling, 1980, p. 132).

Preference of the contamination family as indicated by the numerical results should be preference of $\Delta(g)$ in the preceding approximation. Therefore we compare both approximations with the true power at alternatives g_n , approaching the null hypothesis (not necessarily in a contiguous way). The following theorem clarifies the different structure of the error term of the two asymptotic approaches and shows why the contamination model is preferable. The proof is based on the inequalities given in Theorem 1.1. and 1.2.

Theorem 2.1. Let $k \geq 3$. Let $\{s_n\}$, $s_n > 0$, be a nondecreasing sequence, and let

$$G(s_n) = \left\{ g : \max_i \left| \int_{A_i} g \, d\lambda - \int_{A_i} h \, d\lambda \right| \leq s_n n^{-1/2} \right\}. \quad (2.4)$$

(i) As $n \rightarrow \infty$,

$$\sup_{g \in G(s_n)} \sup_{c > 0} \left| P_g(X^2 > c) - \Pr(\chi_{k-1}^2(n \Delta(g)) > c) \right| = O(s_n n^{-1/2}). \quad (2.5)$$

(ii) There exists a positive constant c_1 such that for all $g \in G(s_n)$ and n ,

$$\sup_{c > 0} \left| P_g(X^2 > c) - \Pr(\chi_{k-1}^2(n \tilde{\Delta}(g)) > c) \right| \leq c_1 s_n n^{-1/2} + (2\pi)^{-1/2} \omega_n(g) n^{1/2}, \quad (2.6)$$

where $\omega_n(g) = |\Delta(g)^{1/2} - \tilde{\Delta}(g)^{1/2}|$, implying that as $n \rightarrow \infty$,

$$\sup_{g \in G(s_n)} \sup_{c > 0} \left| P_g(X^2 > c) - \Pr(\chi_{k-1}^2(n \tilde{\Delta}(g)) > c) \right| = O(s_n n^{-1/2} + \omega_n n^{1/2}) \quad (2.7)$$

with $\omega_n = \sup_{g \in G(s_n)} \omega_n(g)$. Moreover, (2.6) is sharp in the sense that if $g_n \in G(s_n)$,

$$\begin{aligned} \liminf_{n \rightarrow \infty} \max(n \Delta(g_n), n \tilde{\Delta}(g_n)) &> 0, \\ n^{1/2} \omega_n(g_n) &= O(1) \quad \text{and} \quad \lim_{n \rightarrow \infty} n^{1/2} \omega_n(g_n) / (s_n n^{-1/2}) = \infty, \end{aligned} \quad (2.8)$$

then there exists a positive constant c_2 such that for all n ,

$$\sup_{c > 0} \left| P_{g_n}(X^2 > c) - \Pr(\chi_{k-1}^2(n \tilde{\Delta}(g_n)) > c) \right| \geq c_2 n^{1/2} \omega_n(g_n). \quad (2.9)$$

Proof. (i) This is a restatement of Theorem 2.1(iii) with $\lambda = 1$ in Drost et al. (1989b).

(ii) Combination of (2.5) and (1.1) yields (2.6). By (2.5) and (1.2) (2.9) follows, thus proving the sharpness of (2.6). \square

Remark 2.1. The error bounds in the theorem are uniform in the critical values c , so the approximations hold uniformly for all significance levels α_n .

As is seen from the theorem the χ^2 -approximation based on the exponential family approach is inferior to the χ^2 -approximation based on the contamination family approach if $n^{1/2} \omega_n(g_n)$ is of larger order than $s_n n^{-1/2}$. (Otherwise both approximations yield the same error bound.) Next it will be shown that typically $n^{1/2} \omega_n(g_n)$ is indeed of larger order than $s_n n^{-1/2}$ as soon as $s_n \rightarrow \infty$. So for contiguous alternatives the same error bound is obtained, but for more distant alternatives the χ^2 -approximation based on the contamination family approach is better. In a way this corresponds to the numerical results in Kallenberg et al. (1985). For alternatives with heavy tails there were conflicting theoretical results with high power in the contamination family approach and low power in the exponential family approach. The numerical results were indeed in favour of the contamination family approach. This agrees with the preceding theorem in the sense that for more distant alternatives there is a discrepancy between both approximations in favour of the contamination family approach.

We write $z_n = (g_n/h) - 1$. To show that typically $n^{1/2}\omega_n(g_n)$ is of larger order than $s_n n^{-1/2}$ as $s_n \rightarrow \infty$ we consider alternatives g_n such that

$$\sup_{y \in S} \left| \frac{g_n(y) - h(y)}{h(y)} \right| \leq s_n n^{-1/2}, \quad \lim_{n \rightarrow \infty} s_n n^{-1/2} = 0$$

and

$$\liminf_{n \rightarrow \infty} \left| \sum_{i=1}^k \pi_i^{-1} \int_{A_i} h z_n d\lambda \int_{A_i} h z_n^2 d\lambda \Delta(g_n)^{-1/2} s_n^{-2} n \right| > 0. \quad (2.10)$$

Note that (2.10) implies that $g_n \in G(s_n)$. By Taylor expansion we have

$$\begin{aligned} \int_{A_i} h \log(g_n/h) d\lambda &= \int_{A_i} h \log(1 + z_n) d\lambda \\ &= \int_{A_i} h z_n d\lambda - \frac{1}{2} \int_{A_i} h z_n^2 d\lambda + O(s_n^3 n^{-3/2}) \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Hence, using $\int h z_n d\lambda = 0$,

$$\begin{aligned} \tilde{\Delta}(g_n) &= \sum_{i=1}^k \pi_i^{-1} \left\{ \int_{A_i} h z_n d\lambda - \frac{1}{2} \int_{A_i} h z_n^2 d\lambda + \frac{1}{2} \pi_i \int_{A_i} h z_n^2 d\lambda + O(s_n^3 n^{-3/2}) \right\}^2 \\ &= \Delta(g_n) - \sum_{i=1}^k \pi_i^{-1} \int_{A_i} h z_n d\lambda \int_{A_i} h z_n^2 d\lambda + O(s_n^4 n^{-2}) \end{aligned}$$

and therefore

$$\begin{aligned} n^{1/2}\omega_n(g_n) &= n^{1/2} \left| \Delta(g_n)^{1/2} - \tilde{\Delta}(g_n)^{1/2} \right| \\ &= \frac{1}{2} n^{1/2} \left| \sum_{i=1}^k \pi_i^{-1} \int_{A_i} h z_n d\lambda \int_{A_i} h z_n^2 d\lambda \Delta(g_n)^{-1/2} + O(s_n^3 n^{-3/2}) \right|. \end{aligned}$$

By (2.10) the last expression is of *exact* order $s_n^2 n^{-1/2}$, implying that $\omega_n(g_n)$ is of larger order than $s_n n^{-1/2}$ if $s_n \rightarrow \infty$ as $n \rightarrow \infty$.

This completes the proof of the following theorem, expressing the preference for the contamination family approach.

Theorem 2.2. Let $k \geq 3$. Let $\{s_n\}$, $s_n > 0$, be a nondecreasing sequence, and let $G(s_n)$ be defined by (2.4). Consider a sequence of alternatives $\{g_n\}$, satisfying (2.8) and (2.10) with $\lim_{n \rightarrow \infty} s_n = \infty$. Then

$$\lim_{n \rightarrow \infty} \frac{\sup_{c > 0} |P_{g_n}(X^2 > c) - \Pr(\chi_{k-1}^2(n \Delta(g_n)) > c)|}{\sup_{c > 0} |P_{g_n}(X^2 > c) - \Pr(\chi_{k-1}^2(n \tilde{\Delta}(g_n)) > c)|} = 0. \quad \square$$

Appendix

Lemma A.1. Let

$$\tilde{\pi}_i(\theta_n) = \int_{A_i} \tilde{g}_{\theta_n} d\lambda$$

and $\tilde{\Delta}$ as defined in (2.3). Assume that $\lim_{n \rightarrow \infty} n\theta_n^2$ finitely exists and $\int h \log(g/h) d\lambda > -\infty$. Then

$$\lim_{n \rightarrow \infty} n \sum_{i=1}^k \{ \tilde{\pi}_i(\theta_n) - \pi_i \}^2 \pi_i^{-1} = \lim_{n \rightarrow \infty} n\theta_n^2 \tilde{\Delta}. \quad (\text{A.1})$$

Proof. It suffices to show that

$$\lim_{n \rightarrow \infty} \frac{\tilde{\pi}_i(\theta_n) - \pi_i}{\theta_n} = \int_{A_i} f h \, d\lambda$$

with

$$f = \log(g/h) - \int h \log(g/h) \, d\lambda.$$

Note that $\int h \log(g/h) \, d\lambda > -\infty$ implies $\int h |\log(g/h)| \, d\lambda < \infty$. For all $v > 0$ and $0 < \theta < 1$ we have

$$|(v^\theta - 1)/\theta| \leq v + |\log v|.$$

Put $v = g/h$. By dominated convergence it follows that

$$\int (1 - v^\theta) \theta^{-1} h \, d\lambda \rightarrow - \int (\log v) h \, d\lambda \quad \text{as } \theta \downarrow 0.$$

Hence, the normalizing constant $c(\theta)$ of the exponential family (cf. (2.2)), satisfies, as $\theta \downarrow 0$,

$$\theta^{-1} \{c(\theta) - 1\} = \int \theta^{-1} (1 - v^\theta) h \, d\lambda / \int v^\theta h \, d\lambda \rightarrow - \int (\log v) h \, d\lambda.$$

For sufficiently large n we have

$$\left| \frac{\tilde{g}_{\theta_n} - h}{\theta_n} \right| \leq c(\theta_n) \left| \frac{v^{\theta_n} - 1}{\theta_n} \right| h + \left| \frac{c(\theta_n) - 1}{\theta_n} \right| h \leq 2 \{v + |\log v|\} h + \left\{ -2 \int (\log v) h \, d\lambda \right\} h$$

and therefore by dominated convergence

$$\lim_{n \rightarrow \infty} \frac{\tilde{\pi}_i(\theta_n) - \pi_i}{\theta_n} = \lim_{n \rightarrow \infty} \int_{A_i} \frac{\tilde{g}_{\theta_n} - h}{\theta_n} \, d\lambda = \int_{A_i} f h \, d\lambda. \quad \square$$

Acknowledgment

The author would like to thank the referee for suggesting an improvement of Theorem 1.1.

References

- Brown, L.D. (1971), Non-local asymptotic optimality of appropriate likelihood ratio tests, *Ann. Math. Statist.* **42**, 1206–1240.
- Drost, F.C., W.C.M. Kallenberg, D.S. Moore and J. Oosterhoff (1989a), Asymptotic error bounds for power approximations to multinomial tests of fit, in: L.J. Gleser et al., eds., *Contributions to Probability and Statistics: Essays in honor of Ingram Olkin* (Springer, New York) pp. 429–446.
- Drost, F.C., W.C.M. Kallenberg, D.S. Moore and J. Oosterhoff (1989b), Power approximations to multinomial tests of fit, *J. Amer. Statist. Assoc.* **84**, 130–141.
- Gvanceladze, L.G. and D.M. Chibisov (1979), On tests of fit based on grouped data, in: J. Jurečková, ed., *Contribution to Statistics, J. Hájek Memorial Volume* (Academia, Prague) pp. 79–89.
- Kallenberg, W.C.M., J. Oosterhoff and B.F. Schriever (1985), The number of classes in chi-squared goodness-of-fit tests, *J. Amer. Statist. Assoc.* **80**, 959–968.
- Kester, A.D.M. and W.C.M. Kallenberg (1986), Large deviations of estimators, *Ann. Statist.* **14**, 648–664.
- Quine, M.P. and J. Robinson (1985), Efficiencies of chi-square and likelihood ratio goodness-of-fit tests, *Ann. Statist.* **13**, 727–742.
- Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics* (Wiley, New York).